# Why Invest There?

FDI may follow incentives,
but incoming firms are also attracted to the
presence of industry clusters and the benefits of colocation

KELLEY SCHOOL OF BUSINESS
INDIANA UNIVERSITY
Indiana Business Research Center

INDIANA UNIVERSITY
CIBER

# Why Invest There?

**FDI may follow incentives, but incoming firms are also attracted to the presence of industry clusters and the benefits of colocation**

**September 2018**

Timothy F. Slaper, Ph.D.

Director of Economic Analysis,
Indiana Business Research Center,
Indiana University Kelley School of Business

Ping Zheng, Ph.D.

Economic Research Analyst,
Indiana Business Research Center,
Indiana University Kelley School of Business

Prepared for:
Center for International Business Education and Research

Prepared by:
Indiana Business Research Center
Kelley School of Business
Indiana University

www.ibrc.indiana.edu

# Table of Contents

Executive Summary ..... 1
Introduction ..... 2
The Theory behind Clusters, Diversification and Colocation Externalities ..... 3
Data Sources and Measures ..... 6
Analysis and Results ..... 11
Discussion and Conclusion ..... 14
References ..... 15
Appendix ..... 16

## Index of Figures

**Figure 1:** Number of FDI Projects, 2007-2015 ..... 8
**Figure 2:** Anticipated FDI Employment Gains, 2007-2015 ..... 9
**Figure 3:** Planned FDI Investments (in millions of dollars), 2007-2015 ..... 10

## Index of Tables

**Table 1:** Source of Data Used in the Analysis ..... 6
**Table A1:** Summary of Empirical Results of Pooled Sample ..... 16
**Table A2:** Summary of Empirical Results of Panel Models for FDI Employment ..... 18
**Table A3:** Variable Names and Definitions ..... 20

A more complete presentation of the research, analysis and results that contributed to the content of this report can be found at:
https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3228722

# Executive Summary

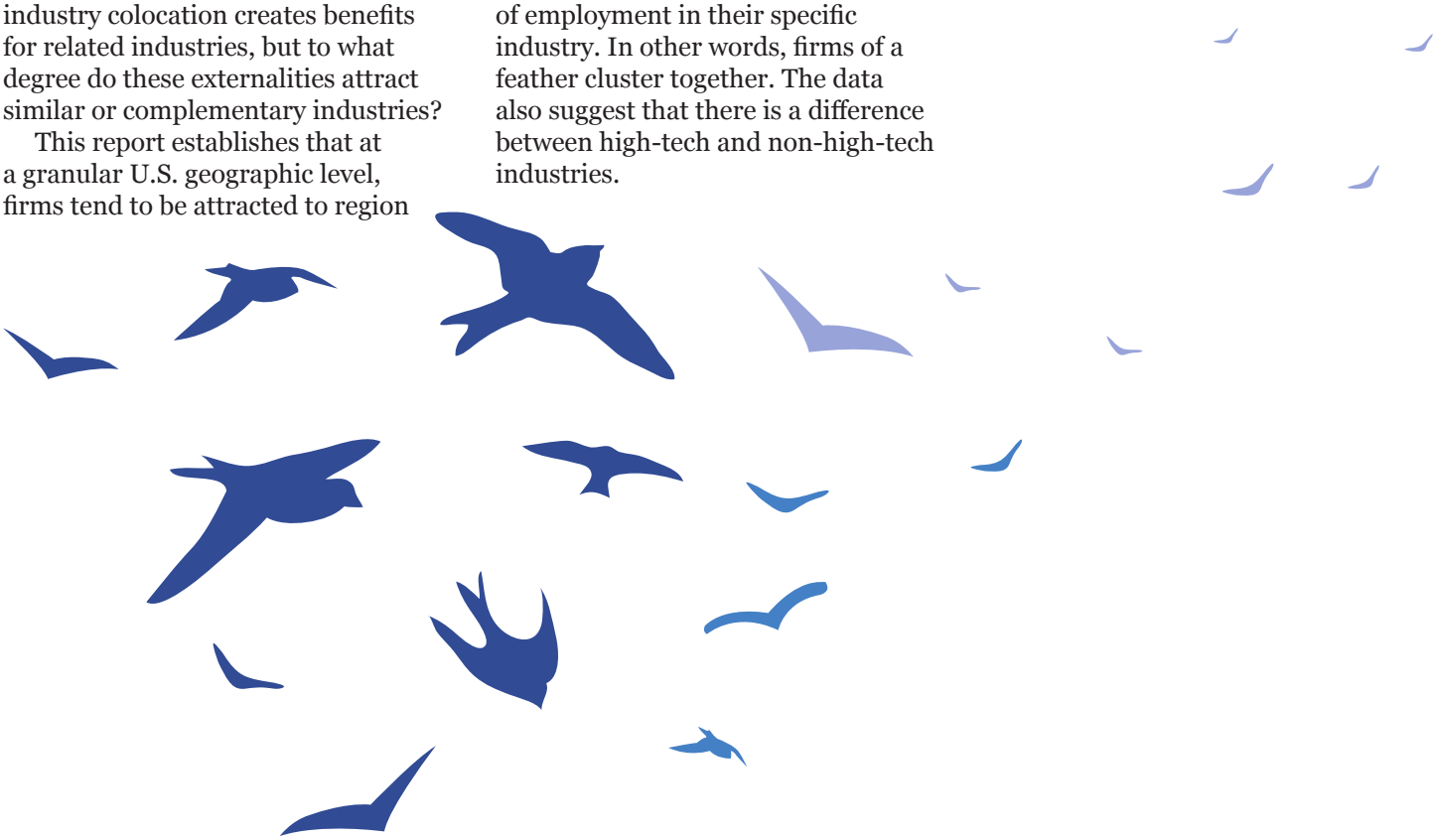
There is an expression: Birds of a feather flock together. This observation and expression has something of an equivalence in economics—agglomeration of similar companies. That is, in economics, there are competitive and productivity benefits associated with the concentration and colocation of related industries—often called industry clusters. Why birds of a feather may flock and colocate together is not of our concern here, but we are interested in whether and the extent to which firms in certain industries choose to locate together.

It turns out that U.S.-based research is sparse on whether regions with specialized industry clusters—that is, industries flocking together if the reader will allow—magnetically attract investment from firms outside the region. According to theory, industry colocation creates benefits for related industries, but to what degree do these externalities attract similar or complementary industries?

This report establishes that at a granular U.S. geographic level, firms tend to be attracted to region (i.e., counties) that have an absolute concentration of employment in their industry cluster. Second, the authors also find that high-tech industries have a different foreign direct investment (FDI) attraction profile than non-high-tech industry clusters, an important consideration for economic development practitioners to consider as they create their development strategies. Third, several regional characteristics that are considered important by site selectors—those informing the FDI location decisions—are more salient than other regional characteristics and attributes.

In this report, we use greenfield foreign direct investment data at the U.S. county level and find that firms are more likely to invest in new or expanded facilities in regions that have a high absolute concentration of employment in their specific industry. In other words, firms of a feather cluster together. The data also suggest that there is a difference between high-tech and non-high-tech industries.

We also find that that several regional characteristics such as the availability of labor is an important consideration, together with state-level characteristics such as lower electricity costs and good state governance. These results are largely similar and robust across statistical methods, irrespective of whether the variable of interest (the dependent variable) is receiving FDI of any kind, the number of projects or the number of jobs associated with the FDI.



# Introduction

The U.S. literature on the benefits of firm colocation has largely been silent on whether specialized or diversified production structures attract external economic investment. Much has been written on the degree to which colocation creates benefits to like kinds of companies sharing geographic proximity. Yet, empirical analysis of the extent to which the externalities of colocation attract similar or complementary industries has largely gone unstudied.

In this report, we address whether, and the degree to which, colocation externalities provide something of a magnetic attraction and influence a firm's decision to invest in plant and equipment in one place versus another. Using greenfield foreign direct investment for the U.S., we conclude that firms are more likely to invest in new or expanded facilities in regions that have a relatively high concentration of employment in their particular industry or among complementary industries.

The contribution to the literature is threefold. First, the report establishes that at a granular U.S. geographic level, firms tend to be attracted to regions—counties—that have an absolute concentration of employment in their industry cluster. Second, we also find that high-tech industries have a different FDI attraction profile than non-high-tech industry clusters, an important consideration for economic development practitioners to consider as they create their development strategies. Third, we find that several regional characteristics that are considered important by site selectors—those informing the FDI location decisions—are more salient than other regional characteristics and attributes.

The report first presents an overview of the theory associated with the arguments for industry specialization, diversification and location externalities. In Section 2, we describe the data and measures. Our empirical method and results are presented in Section 3. We conclude with a brief discussion and conclusion in the last section.

# The Theory behind Clusters, Diversification and Colocation Externalities

Are industry clusters the same as industry sectors? Industry clusters—also known as business clusters or competitive clusters—are groups of similar or related firms that play an important role in the success of regional economic development. Clusters share a common market, facilitate the exchange of suppliers, and develop localized worker skills and know-how. An industry cluster represents a broadly defined industry from suppliers to end producers, which is different from the classically defined industry sectors that are organized according to production technologies.

Clusters, like cheese-making in Vermont or tech-companies in California's Silicon Valley in California, can strengthen competitiveness as they increase productivity, simulate innovative new partnerships, and present opportunities for entrepreneurial activity. Suppose you would like to expand your surfboard business, which is a better choice, Hawaii or Colorado? It isn't a stretch to argue that relocation decisions are made due to the presence of strong industry cluster.

Industry clusters are commonly used to measure what kind of firms locate in close proximity. The empirical evidence indicates that there are competitive and productivity benefits associated with the concentration and colocation of related industries and in this report, these benefits are called colocation externalities.

It has sometimes been said that clusters form "because there is something in the air." That something is beneficial externalities associated with similar or related firms sharing geographic proximity. The benefits of proximity, or colocation, are also referred to as "localization externalities." Long-established firms benefit as they are effectively forced to become more productive by competitive pressures. New firms—start-ups—can also take advantage of a well-developed regional labor force and supply chain. One might say that these firms grow based on the resources, labor and know-how in the regions, as well as technology from outside the region—combined with increasing demand for the cluster's goods and services from outside the region. The industrial metabolism of clusters in the region benefits both established firms and local start-up firms.

On the other hand, clusters can also grow "magnetically," that is, a region can attract firms to take advantage of that region's competitive advantage in resources, supply networks and human talent. There might be significant benefits to close geographic proximity for young or mature firms to move into the region in order to take advantage of the colocation externalities. Greenfield foreign direct investment (FDI) is an example of magnetic growth.

Either way, via metabolic growth or magnetic growth, one can understand why the concept and presence of industry clusters may attract the attention of those advocating for a region or state's economic growth. Indeed, the importance of industrial clusters to boost regional economic development has widely gained scholars' attention, most notably, the Harvard Professor Michael Porter (1998 and 2003). Much of the empirical work focuses on the benefits of clusters on industrial employment, innovation and productivity, but less systematic empirical attention has been paid to identifying strong regional clusters and the regional characteristics that attend cluster formation and growth.

For this reason, the authors of this report wanted to empirically confirm whether strong, established, growing clusters tended to attract incoming firms in the form of any "foreign" direct investment? ("Foreign" here is any investment from outside the region regardless of international and national.) But before we present the empirical results, a short discussion on the theoretical basis of clusters may help to clear up any possible confusion between, say, cluster-based economic development and specialization-based economic development strategy, or other parallel concepts.


Industry cluster strength can be viewed as the relative concentration of an industry cluster, without regard to the balance or concentration of industries within that cluster, or as the absolute concentration of employment

in a cluster. Economic development practitioners and policymakers often use the concept and measure of a location quotient (LQ). An LQ simply compares a proportion of industry employment of a region against some other benchmark like the national average proportion of employment of an industry. If one hears that Indiana is the most manufacturing-intensive state in the union, it isn't because it has more workers in manufacturing than, say, Texas, the claim is based on an LQ. Indiana has more workers in manufacturing as a proportion of its workforce than any other state (with Wisconsin close behind). The LQ isn't the last word, or metric, in terms of specialization. Many researchers make the case that specialization (or concentration) is better measured based on absolute size rather than location quotient because, depending on the industry, a region needs a critical mass of employment to be considered specialized in an industry.

Industry cluster strength aligns with the notion of related variety. At first blush, related variety may seem like something of a head scratcher. An example may help. Consider Orlando, Florida. There are several amusement parks. There aren't three Disney Worlds or three Universal Studios in Orlando. A variety of amusement parks exist for the under-stimulated public to enjoy.

In a similar fashion, there is related variety among firms and industries. For example, one can categorize industries based on their technological and material requirements. Orlando might also be a good example of how related variety benefits competitors. While Universal Studios and Walt Disney World may compete for patrons, the presence of several similar venues may make use of a network of local food vendors, security services, landscaping and transportation companies, as well as a labor force attuned to the safety and mechanical requirements of running technologically advanced thrill rides. The colocation benefits of such related variety were first conceptualized by famed-economic Alfred Marshall over a century ago (1966). As a result, these byproducts of related variety are often referred to as MAR externalities. (The M in MAR is for Marshall.) MAR externalities are within related industries, usually broadly defined, but in this case MAR externalities would be in evidence within a cluster. "MAR cluster" is hereafter the term used for related industry concentration and its attendant benefits.

While not the polar opposite of specialization, diversity is in contrast to specialization. That is, for example, if one's stock portfolio consists of one firm that would be the polar opposite of diversity, but one may have a stock portfolio with all technology stocks. While specialized in technology firms, there would be some diversity of firms within that category of stocks. This latter example would be akin to a related variety of stocks. An example of unrelated variety of stocks may be one consisting of Nike, Kroger, Hilton and Sprint. Four firms in four very different markets.

In our case, a diversity of industries has both an industry cluster dimension, that is, the portfolio of industry clusters in a region, as well as the portfolio of related industries within an industry cluster. Both are akin to the diversification of stocks in a portfolio, but one is regional diversification of industry clusters, while the other is within-industry cluster diversification of industries that, by classification, are aggregated into an industry cluster. In other words, cluster diversity used here is not a measure of how, and in what ways, unrelated clusters are different from each other; rather diversity is more synonymous with industry balance within a region or within an industry cluster.

These concepts and definitions place the investigation into context. That is, investigating the role of magnetic cluster growth in U.S. regions, in the form of greenfield and expansionary investment flows, i.e., FDI. To what degree, then, do colocation externalities motivate a firm's decision to move into a region, examining whether a high concentration of related industries, or strong clusters, tend to attract additional investment inflows and thus additional employment within that cluster? In addition, we are particularly interested in whether a more diversified, or balanced, set of industries within a strong, or highly concentrated cluster—an MAR cluster—tends to attract new greenfield investment or additional

expansionary investment among firms already operating in the region.

The opposite of magnetic cluster growth may be thought of as the "throw-a-dart" approach to site location (Ellison and Glaeser, 1997). Do the location patterns of new investment in plant and equipment, and the concomitant employment, follow a random, throw-a-dart approach or reflect decisions that may be motivated by seeking the competitive benefits of industrial colocation? For this reason, we can identify investment moving into a region and assess whether a region has relative strength, or specialization, in the cluster associated with the investment. In addition, we can also assess whether the receiving regional industry cluster has a diverse set of industries or is simply dominated by one or two industries within that industry cluster.

One can also hypothesize about the level of the associated technological sophistication for the new employment. The investment in clusters that are not in the high-technology domain is well in evidence after the Great Recession. Initial analysis also shows that the clusters in the top 10 list in terms of the number of incoming jobs tend to be more balanced (diversified). This may signal the importance of a well-developed labor force, as well as supply chains and material linkages among colocated firms.

Finally, the authors also explored additional dimensions that an investor, or site selector, may consider important to a location decision. Based on a report by the International Economic Development Council, a Washington, DC-based association for advancing regional economic development, the most important factors in business location decisions included infrastructure, workforce characteristics, wages, labor market, demographics, higher education, labor regulations, taxes and incentives (2016). While many of these characteristics are available at the county level, the latter three items are more closely aligned with state policy and practices. If these regional or county characteristics do influence the decision to invest in one region as opposed to another, then these characteristics need to be controlled in the empirical model. For example, to control for state policy effects, we also included state-based proxy data that investors may consider as indicators of good state governance.

# Data Sources and Measures

To examine region-cluster employment growth, we draw on studies of regional economic growth as a function of the level of economic activity and attributes of the region. The econometric model regresses announced investment on plant, equipment and employment on a number of factors that characterize cluster strength. An entropy index is also used to assess the diversity or evenness of industry clusters in a region, as well as diversity of industries within an industry cluster. The list of variables used is shown in **Table 1**.

We used employment by industry data from QCEW-complete employment estimates, aggregating into industry clusters based upon definitions from the U.S. Cluster Mapping Project (CMP). Thus, all the industry data are bundled into "industry clusters," of which there are 70. The proprietary data set for greenfield and expansion employment and the number of investment projects associated with investment announcements is from fDiMarkets.

There are several potential weakness associated with the FDI announcement data. One, the jobs realized once the plant and equipment are in operation may be different than the number of jobs reported in the press releases. Two, there is no way to verify how many new, incoming magnetic jobs, were created, because of the disclosure constraints associated with record-level QCEW establishment data. In other words, one cannot link an FDI announcement record in 2012 with subsequent establishment data. Three, there is no fixed time between an FDI press release and realized jobs. The latter can vary greatly depending on the industry, the scale of investment, market demand conditions for the firms, etc. That said, firms have been known to spend several years and millions of dollars in site selection and negotiating with local and state officials before making an announcement; thus, we consider the FDI announcements as an appropriate signal for a region's relative attractiveness in terms of agglomeration externalities.

The unit of analysis is industry clusters at the U.S. county level. While some FDI projects could be considered "local" (for example, real

**Table 1: Source of Data Used in the Analysis**

| Data | Source | Years |
|---|---|---|
| **County-Level Data** | | |
| Employment by industry | QCEW-complete employment estimates (Indiana Business Research Center) | 2004-2015 |
| Greenfield and expansion employment | fDiMarkets | 2007-2015 |
| Number of investment projects associated with investment announcements | fDiMarkets | 2007-2015 |
| Educational attainment (percent of population with less than a high school diploma, some college, and bachelor's degree or above) | American Community Survey (U.S. Census Bureau) | 2004-2015 |
| Percent of population in prime working ages (ages 25-44) | American Community Survey (U.S. Census Bureau) | 2004-2015 |
| Average travel time to workplace | American Community Survey (U.S. Census Bureau) | 2007-2015 |
| Number of STEM graduates | IPEDS (National Center for Education Statistics) | 2004-2015 |
| Unemployment rate | U.S. Bureau of Labor Statistics | 2004-2015 |
| Average hourly wage in manufacturing | U.S. Bureau of Labor Statistics | 2004-2015 |
| Cost of living index | Council for Community and Economic Research | 2016 |
| Transportation cost of living | Council for Community and Economic Research | 2016 |
| Interstate lane miles | U.S. Department of Transportation | 2015 |
| Cost of electricity for industrial use | U.S. Department of Energy | 2004-2015 |
| University knowledge spillovers | Indiana Business Research Center | 2010-2015 |
| **State-Level Data** | | |
| State business tax climate index | Tax Foundation | 2015 |
| State and local taxes per capita | Tax Foundation | 2012 |
| State and local tax as a percent of state income | Tax Foundation | 2012 |
| Percentage of public pension plans that are funded | Tax Foundation | 2014 |
| State credit ratings | Standard & Poors | 2004-2015 |
| Venture capital | Thompson Reuters | 2004-2015 |

Note: All the industry data are bundled into 70 "industry clusters."
Source: Indiana Business Research Center

estate development or consumer banking branches)—in contrast to "traded" industries for which the market generally extends beyond the region—the vast majority of FDI announcements are for traded industries and we consider only traded industry clusters. We collected nine years of FDI data (2007 to 2015), grouping them into three three-year time periods since FDI data by county tends to be sparse. Of the 30,774 FDI announced events in the U.S. over this time period, 20,632 were in the relevant traded industries.

There are 243,698 county-by-industry-cluster observations for the 2007 to 2015 time period, implying that the average county has about 27 traded industry clusters, based on the QCEW-complete data. Around 40 percent of those county-industry clusters are sufficiently concentrated to be considered an MAR cluster in the "cluster development strategy" sense (i.e., an above national average concentration of related industries that tend to benefit from economies of agglomeration). In other words, there are about 40 percent of county-by-industry-cluster observations that are, on a relative basis, "true" MAR clusters with LQs above 1.2. Of these MAR clusters, around 14 percent are high-tech industries.

County-by-industry-clusters can have multiple projects or FDI attraction events over the time period. As a result, the number of county-by-industry-clusters that recorded FDI employment announcements is whittled down to 8,194. The average number of jobs per FDI announcement is 190, but each FDI project or event can range from one new job to over 8,000. For an example of the latter, the IT sector in Travis County, Texas, as in Austin, attracted 8,000 new workers based on an FDI announcement.

While the main data source for the explanatory variables is QCEW data by industry, how these data are operationalized to provide measures of regional agglomeration and industry structure warrants discussion.

There is a wide variation in terms for cluster concentration/specialization. For example, farming regions and regions endowed with natural resources tend to have very high employment concentration in specialized sectors. On the other hand, high-tech clusters, especially the ones associated with FDI—for example, Travis County, Texas—tend to have a more diversified economy and industrial profile.

We use a common entropy measure—the Shannon Evenness Index—to assess the degree to which a region's industry clusters are even/balanced or uneven. An index value of zero (perfect unevenness) occurs if there is only one industry in the region, whereas 1 denotes a perfect balance among industry clusters.

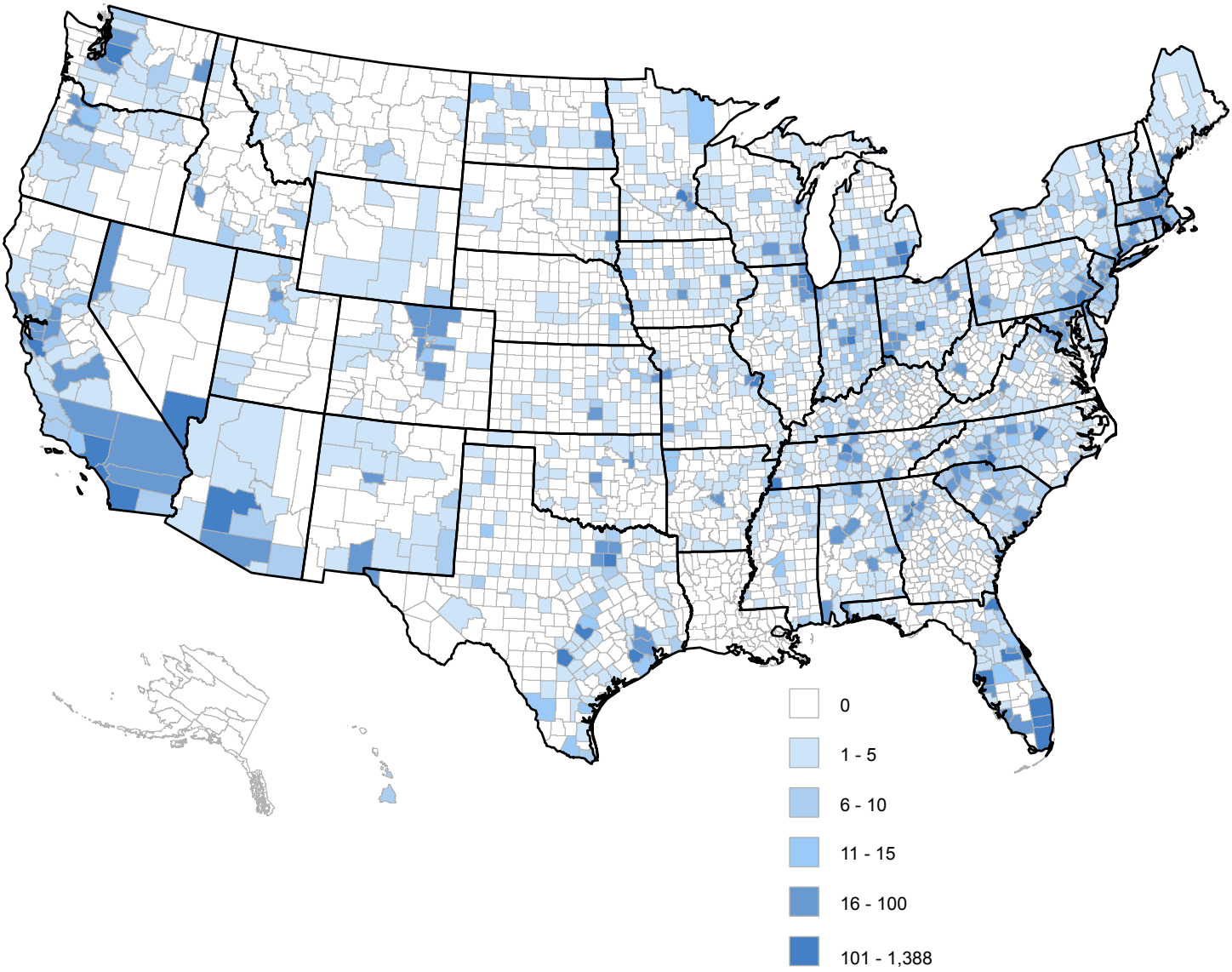
We used two variables for industry cluster strength, or specialization: the value of the location quotient (that measures relative magnitude or degree) and a binary threshold to indicate the presence of an MAR cluster if the location quotient was greater than 1.2.

The three measures discussed above are related to how even/balanced industry clusters are among themselves. These are measures "outside" a particular industry cluster. The following two measures address the balance, evenness or industry specialization within a cluster. As discussed above, the concern is the degree to which an industry cluster in a specific region has the same relative concentration of industries as the nation. Does an industry cluster that is dominated by one particular industry in that cluster yield the same magnetic attraction as an industry cluster that is more balanced? Does a wider complement of industries influence investment decisions?

Finally, in order to provide the reader a sense of where the FDI events are occurring across the country, as well as a sense of the relative importance in a region, **Figures 1** through **3** present county-level maps of the U.S.


Dell main headquarters in Round Rock, Texas. Dell continues to be one of the largest employers in the Austin MSA, as well as 3M, Apple Inc., Hewlett-Packard, Google, Facebook, AMD, Applied Materials, Cirrus Logic, Cisco Systems, eBay/PayPal, Bioware, Blizzard Entertainment, Hoover's, Intel Corporation, National Instruments, Samsung Group, Silicon Laboratories, Oracle Corporation, Hostgator, and United Devices.

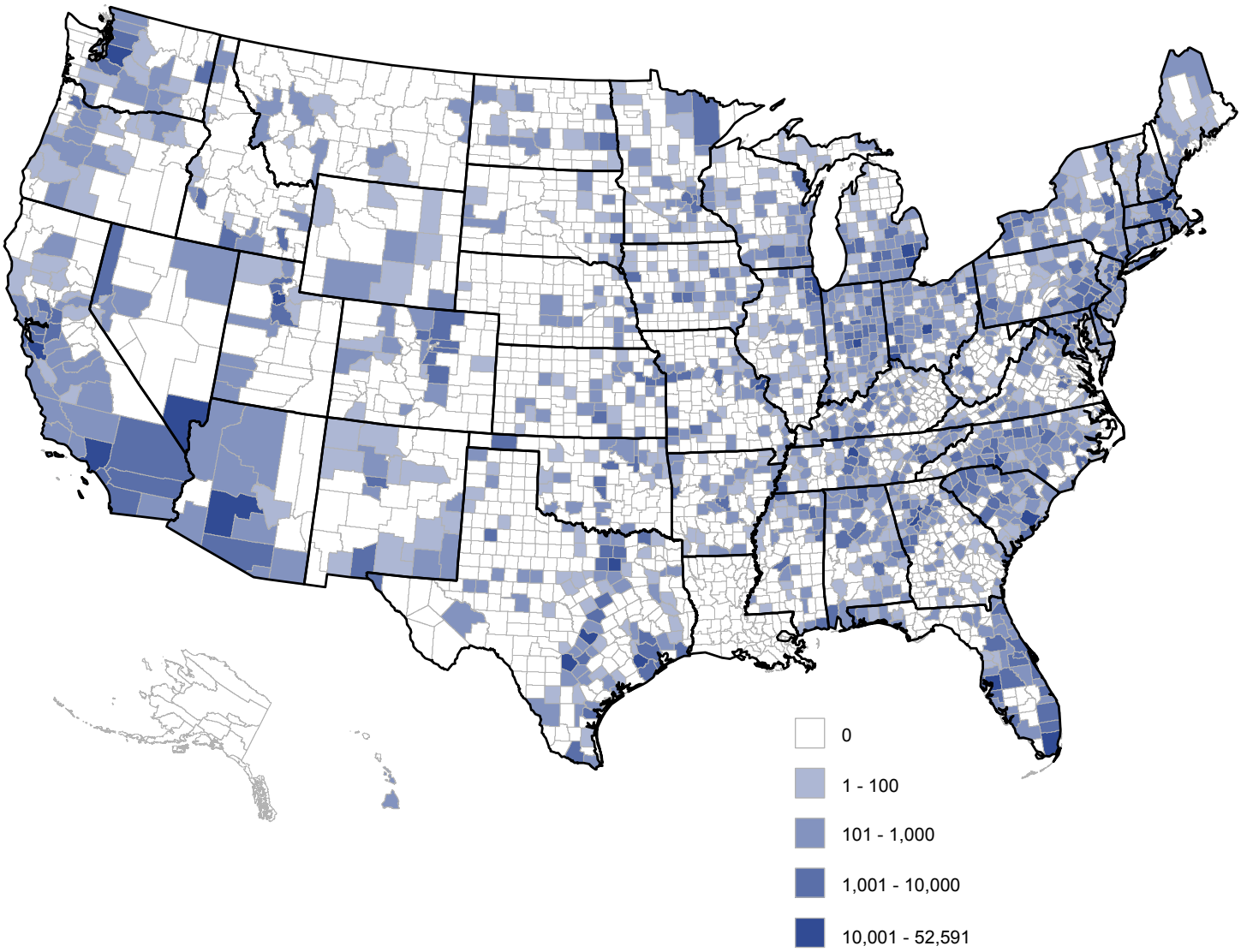
**Figure 1: Number of FDI Projects, 2007-2015**




Source: Indiana Business Research Center, using fDiMarkets data

> **Whether site selectors or corporate decision makers are aware of it or not, the location of FDI projects align with the benefits of highly concentrated clusters, which magnetically attract incoming investment into counties.**

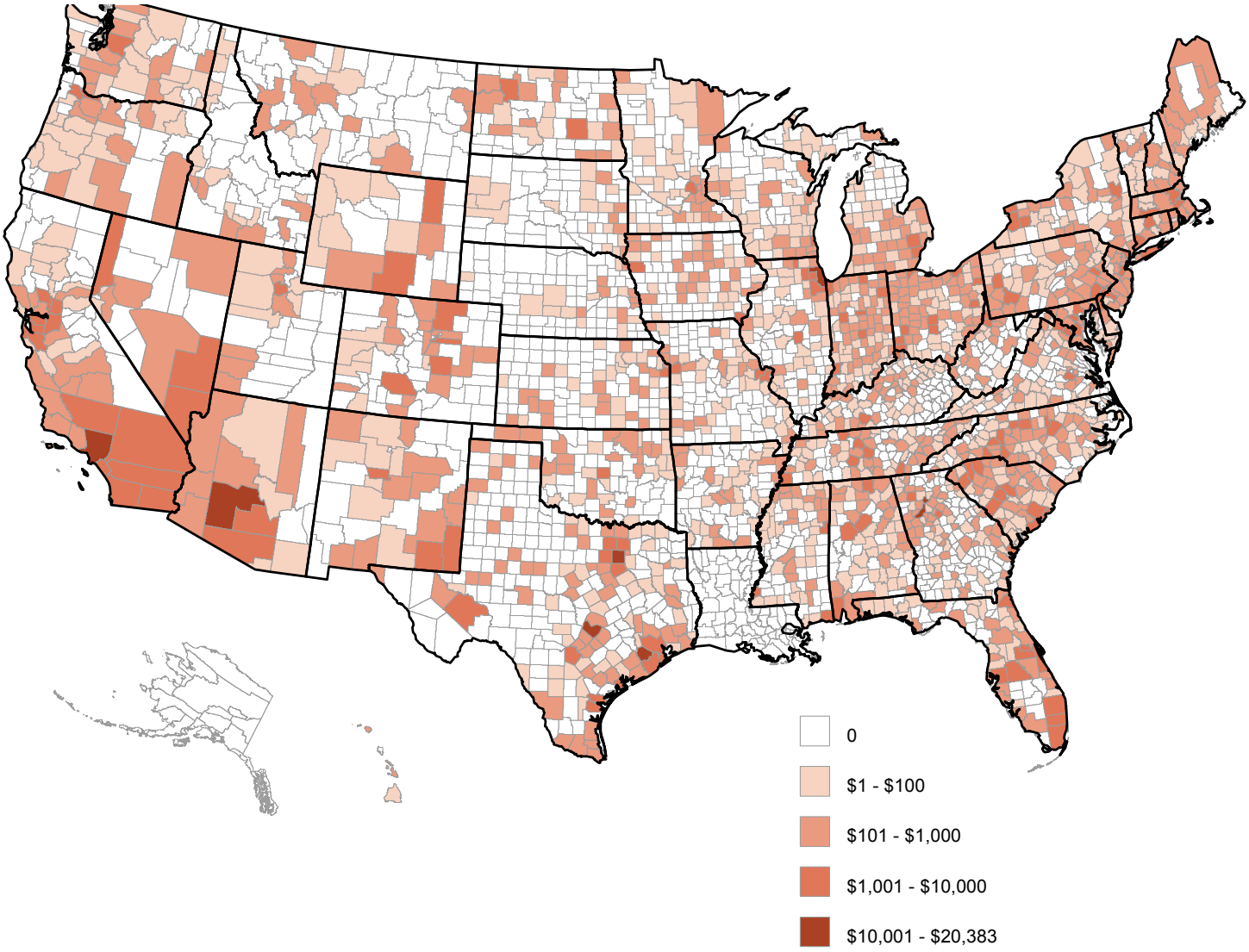
**Figure 2: Anticipated FDI Employment Gains, 2007-2015**




Source: Indiana Business Research Center, using fDiMarkets data

> **Investment in a county as measured by employment growth attributed to greenfield FDI is strongly positively associated with the absolute size of clusters—the bigger a cluster, the greater the magnetic attraction of FDI-related employment.**

**Figure 3: Planned FDI Investments (in millions of dollars), 2007-2015**




Source: Indiana Business Research Center, using fDiMarkets data

# Analysis and Results

We explored several modeling strategies, including ordinary least squares, pseudo-panel, logit and negative binomial procedures. The variation explained by the models, the coefficients and significance of the explanatory variables, are all similar, but vary because the different modeling strategies use different dependent variables.

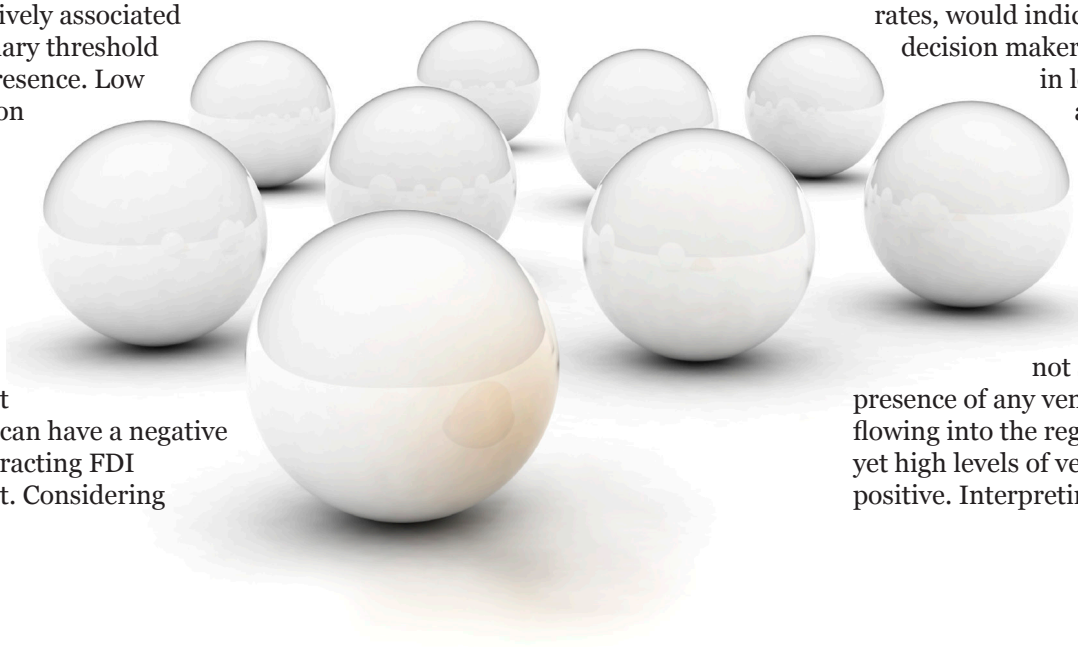
Model 1 (the OLS model) confirms our first hypothesis: Investment in a county as measured by additional employment attributed to greenfield or expansion FDI is strongly positively associated with total cluster employment. The larger the absolute size of an industrial cluster, the greater the magnetic attraction of FDI-related employment. This model suggests that a 1 percent increase in industrial cluster employment is associated with a nearly 30 percent increase in FDI employment. That said, the marginal effect is higher for non-high-tech clusters relative to high-tech clusters. In the appendix, **Table A1** presents the results. (**Table A3** lists variable names and definitions.)

Also in evidence is that MAR clusters—relatively high concentration of an industry cluster—is also positively associated with the binary threshold of cluster presence. Low concentration industry clusters, that is, those with a relatively weak presence of aggregate employment in a region, can have a negative effect on attracting FDI employment. Considering that a vast majority of counties have an LQ of below one (1) for most of a county's industry clusters, this result is not surprising. Moreover, the interaction term for an MAR cluster and industry cluster LQ is also positive, corroborating the evidence that specialization in an industry cluster serves as a magnet for FDI-related employment.

Industry cluster diversity, or evenness across clusters, in a county is negatively associated with an increase in FDI-related employment. That is, those regions that specialize in one or two industry clusters—i.e., less diversified/balanced—tend to receive more FDI-related employment. This effect, however, is mainly driven by high-tech clusters, as shown by the high-tech and Shannon Evenness Index interaction variable (**htflag × sei_clst**). The positive coefficient for the binary, threshold variable for high-tech clusters would indicate that, in general, high-tech industry clusters tend to gain more FDI -related employment than non-high-tech clusters.

Within-industry cluster diversity, a measure of balance for the industries that make up an industry cluster, is negatively associated with FDI. Put another way, specialization within a cluster is positive and statistically significant. However, this effect is offset by the negative effect from the interaction term for high-tech, suggesting that industries in high-tech industry clusters appear to benefit from a more balanced, within cluster, industrial profile for attracting FDI employment. Conversely, industry clusters that are not high-tech would not be penalized for the lack of within-cluster diversity or evenness.

Leaving the magnetic benefits of agglomeration aside for a moment, there are several regional characteristics that may also influence attracting FDI employment. The cost of living (for transportation), for example, changes sign and loses statistical significance once state-level characteristics are considered. The results for level of educational attainment is also ambiguous. The presence of universities graduating STEM degree holders is somewhat positive, indicating that the presence of a robust educational system may positively influence FDI decisions. The presence of prime-working-age adults is strongly positive and statistically significant. That, combined with higher unemployment rates, would indicate that FDI decision makers are interested in locations with abundant labor. (Higher average wage for manufacturing employment, a signal for a tight labor market, was not significant.) The presence of any venture capital flowing into the region was negative, yet high levels of venture capital are positive. Interpreting the worker



commute time (i.e., mean travel time to work) is difficult. Increases in mean travel time may indicate congestion in cities—a negative. On the other hand, long commutes from one rural county to another or from one exurb to another may indicate a degree of labor flexibility and a larger labor shed from which to draw talent.

State-level characteristics that may be relevant to location decisions, such as electricity cost and tax burden measures, were incorporated into the models. It appears that only electricity cost has a strong association with FDI employment—a 1 percent increase in electricity cost is associated with a 26 percent decrease in FDI employment. Of the several measures for good state governance, business conditions and taxes, only state and local taxes appear to have an influence.

The above variables taken together weakly explain the variation of greenfield and expansion FDI-related employment.[1] The overall fit of Model 1 suggests that 16 percent of variation is explained. This result is not entirely different from Model 2 (22 percent) and Model 3 (17 percent). We now turn the attention to those models.

The second model used a logit approach. For Model 2, we estimated the odds of attracting FDI projects—all projects, whether large or small, are counted the same—based on the same variables in Model 1. The signs and statistical significance for the absolute size of cluster employment, within-cluster specialization and high-tech are similar to that of the first model. Where the results diverge are the relative measures of cluster strength, or specialization. Neither the variable indicating the presence of an MAR cluster nor the relative concentration of a cluster are statistically significant. The odds of attracting FDI projects also does not seem to depend on the interplay between the size of the industry cluster and whether the cluster in the region is high-tech or not. For Model 2, educational attainment emerges as a factor in increasing the chances of attracting FDI projects. A higher proportion of the population without a high school degree reduces the chances, while the greater proportion of a region's population with a bachelor's degree or higher increases the odds of attracting FDI projects.

The size of a county as measured by total county employment emerges as increasing the odds of attracting FDI projects, as does relative proximity to universities engaged in STEM-related research and development, as measured by university knowledge spillovers within 50 miles of universities (Zheng and Slaper, 2017). Mean travel time reduces a region's chances, while one measure for infrastructure availability, interstate lane miles per capita, increases a region's attractiveness. Finally, in terms of good governance measures, a state's credit rating positively influences a region's chances for attracting FDI projects.

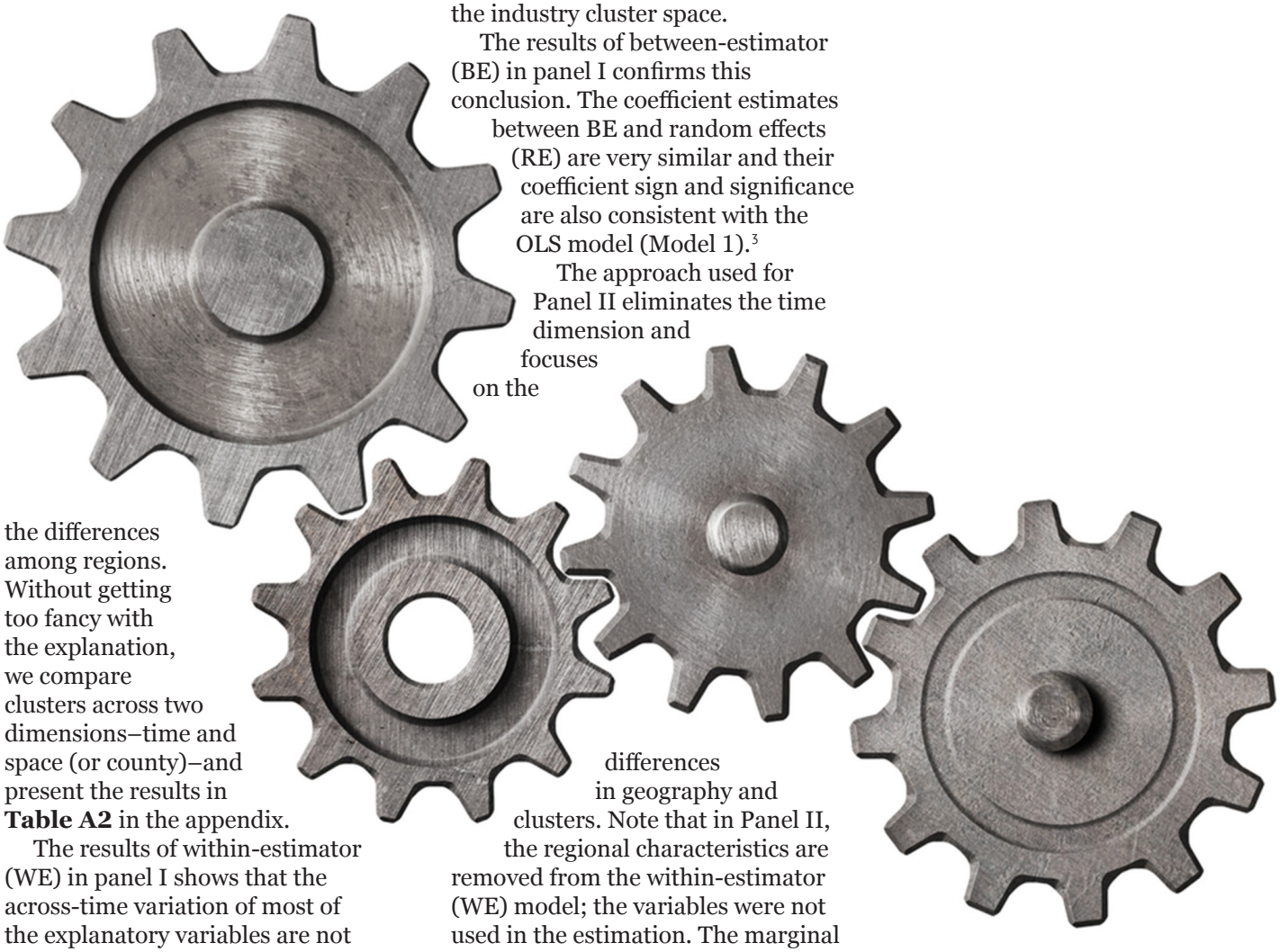
In order to cross-validate the results from Models 1 and 2, Model 3 used a negative binomial model,[2] but as **Table A1** shows, the dependent variable is a count of the number of projects,. In the main, the results were similar to Model 1 for the explanatory variables of particular interest, such as cluster employment and cluster balance. The high-tech and cluster employment interaction term was not statistically significant, however, unlike the OLS employment model. That said, the number of projects and the number of new workers associated with those projects can deviate considerably. In contrast to Model 1, the high-tech and county industry cluster evenness/diversity interaction term lost its significance. The cost of living is positively related to the number of projects, contrary to expectations. The only educational attainment measure with statistical significance is for some college, and then it has a negative sign, contrary to expectations. Then again, if a majority of projects are for manufacturing plants that do not require a highly educated work force, the negative relationship between university education and project counts may not be counterintuitive. The sign for the proximity to university R&D is also contrary to expectations, but may be explained in the same way as education. However, it is difficult to make the latter two results square with the number of STEM degree graduates from institutions in the region as positively related to the number of projects a region attracted. The relationship with interstate lane miles per capita is negative, but this may be indicative of many projects sited in rural counties, far removed from dense development. Rural locations of projects may also help to explain why mean travel time is positively related to the number of FDI projects attracted. As for state characteristics, higher electricity costs are associated with fewer FDI projects, but, contrary to expectations, both state and local taxes per capita and the ratio of state pension plans that are funded are marginally negative (and significant). That said, a good business climate is

[1] While the operative word in the sentence may be "weakly" because the model does not explain a majority of the variation in FDI, when the authors presented the academic paper from which this report was derived, the session participants were impressed with the explained variation. Many of the participants were in the field of economic development and understand the multitude of factors that are in play in site selection, not the least of which are the private negotiations and incentives provided by regional and state governments to secure FDI.

[2] Negative binomial regression is similar to regular multiple regression except that the dependent (Y) variable is an observed count that follows the negative binomial distribution. Thus, the possible values of Y are the nonnegative integers: 0, 1, 2, 3, and so on. Negative binomial regression is often used for over-dispersed count data. That is, if the counts are highly uneven across observations/counties and there are many zero counts, as is the case with 3,110 counties in the U.S., negative binomial regression is often the preferred statistical method.

positively associated with attracting FDI projects.

The last modelling strategy attempts to identify the source of variation that contributes to differences in FDI. For those who are interested in statistics and econometrics, the approach is called a pseudo-panel model. The goal of the analysis is to see if the source of FDI variation is the (possible) change of cluster characteristics over time or if the variation is a result in the differences among regions. Without getting too fancy with the explanation, we compare clusters across two dimensions–time and space (or county)–and present the results in **Table A2** in the appendix.



The results of within-estimator (WE) in panel I shows that the across-time variation of most of the explanatory variables are not significantly associated with FDI employment, with a few exceptions. The sectors that have higher rates of unemployment would attract more FDI employment and so would regions that have more highly educated workers. On the other hand, large high-tech clusters and regions with more prime-working-age populations may reduce the inflow of FDI employment over time. The more compelling results is that the overall fit of the WE model is rather poor. This suggests that time variation is an insignificant source in explaining FDI employment attraction. This leads to the conclusion that the explanatory power of the characteristics associated with attracting FDI employment is better explained by the cross-sectional variation across the industry cluster space.

The results of between-estimator (BE) in panel I confirms this conclusion. The coefficient estimates between BE and random effects (RE) are very similar and their coefficient sign and significance are also consistent with the OLS model (Model 1).[3]

The approach used for Panel II eliminates the time dimension and focuses on the differences in geography and clusters. Note that in Panel II, the regional characteristics are removed from the within-estimator (WE) model; the variables were not used in the estimation. The marginal effects from industry cluster-specific characteristics are much stronger (highly significant and larger) in the between-estimator (BE) model than those from the WE model that eliminates regional characteristics. The results from the BE model are also similar to the BE and RE models in panel I. This leads to the conclusion that industry cluster effects are more dominatent than regional characteristics for FDI location decisions.

[3] For the statistical audience, the over-identifying restriction test shows that the random effects model might be a better choice over fixed effects models. See notes at bottom of Appendix Table 2.

# Discussion and Conclusion

Whether site selectors or corporate decision makers are aware of it or not, the location of FDI projects is not greatly decided by industry cluster specialization, but align with the benefits of MAR clusters, which magnetically attract incoming investment into counties. And whether the industry cluster is high-tech or not may dictate if a region needs to have a balanced, diversified cluster or if specialization in one particular industry within a cluster can be sufficient to attract FDI employment.

Also, certain regional or state characteristics may be important in attracting FDI according to our findings. FDI decisions in educational attainment may relate to the scale and nature of the activity—a small high-tech firm may not consider this an important consideration, but for a 2,000 person manufacturing plant, it is critical. A "flexible" labor market may be an important consideration when considering the scale of a facility: higher unemployment, a large share of prime-working-age population and longer travel time to work may indicate sufficient slack in the labor shed to induce larger facilities to locate in a region. A state's higher credit score may indicate the ability of a state government to negotiate favorable terms with the firm on tax breaks or worker training and retention incentives. Without more granular, case-specific information, these statements are nothing more than hypotheses, but based on our findings, these are credible research paths to explore.

Absent consistent data for site availability or deal-specific details on the incentives—tax reduction benefits or worker hiring and training inducements—to locate in a particular region, the data on FDI location decisions would indicate that colocation economies, lower electricity costs and good state governance conditions drive where greenfield investment and expansions occur.

In summary, we have tested and found valid the claims that the economies of colocation serve as one inducement, among several considerations, for firms to locate in one region as opposed to another. Moreover, these economies of colocation are associated with the absolute scale of the firms in geographic proximity, rather than the relative concentration of those firms within a region. We have also found that a firm within a specifically defined industrial category may be attracted to regions with other firms in close proximity within the same, specific, industrial category. But this depends on the type of industry. For example, high-tech firms appear to seek the presence of a cluster that is internally well-balanced. In contrast, internal cluster balance does not appear to be a concern for non-high-tech firms. Finally, we have found several important regional and state characteristics that appear to have motivated FDI decisions. Electricity costs have a strong association with FDI employment and state and local taxes also appear to have an influence on site location decisions.

**Economic development practitioners and policymakers take note!**

# References


- Ellison, G., & Glaeser, E. L. (1997). Geographic concentration in US manufacturing industries: A dartboard approach. *Journal of Political Economy, 105(5)*, 889-927.
- International Economic Development Council. (2016). *A new standard: Achieving data excellence in economic development*. Washington, DC: Author.
- Marshall, A. (1966). *Principles of economics* (8th ed.). London: Macmillan.
- Porter, M. E. (1998). Clusters and competition: New agendas for companies, governments, and institutions. In M. E. Porter (Ed.), *On competition* (pp. 197–299). Boston, MA: Harvard Business School Press.
- Porter, M. E. (2003). The economic performance of regions. *Regional Studies* 37, 549–578.
- Zheng, P., & Slaper, T. F. (2016). University knowledge spillovers, geographic proximity and innovation: An analysis of patent filings across US counties. [Presentation]. *Paper presented at the 63th Annual North American Meetings of the Regional Science Association International. Minneapolis, MN*. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2857130

# Appendix

### Table A1: Summary of Empirical Results of Pooled Sample

| | FDI employment | FDI received or not | No. of projects |
|---|---|---|---|
| | (Model 1) | (Model 2) | (Model 3) |
| lnclstemp | 0.256*** | 0.434*** | 0.319*** |
| | (0.02) | (0.01) | (0.02) |
| sei_clst | 0.609 | 1.341 | 0.378 |
| | (0.78) | (0.91) | (0.61) |
| lnclst_bal1 | 0.211*** | 0.224*** | 0.080*** |
| | (0.03) | (0.03) | (0.02) |
| htflag | 2.202** | 2.141*** | 1.496* |
| | (0.89) | (0.66) | (0.84) |
| lq_bin | 0.151** | 0.025 | 0.165** |
| | (0.07) | (0.06) | (0.07) |
| lnlq_clst | -0.123*** | -0.017 | -0.096*** |
| | (0.04) | (0.03) | (0.04) |
| lq_bin × lnlq_clst | 0.265*** | | |
| | (0.06) | | |
| htflag × lnclstemp | -0.061*** | -0.007 | -0.004 |
| | (0.02) | (0.02) | (0.01) |
| htflag × sei_clst | -2.107* | -1.444* | -1.765 |
| | (1.15) | (0.83) | (1.08) |
| Htflag × lnclst_bal1 | -0.243*** | -0.380*** | -0.169*** |
| | (0.05) | (0.04) | (0.04) |
| Coli_trsp | 0.001 | -0.001 | 0.013*** |
| | (0.01) | (0.01) | (0.00) |
| nohs | 0.012 | -0.125** | -0.046 |
| | (0.05) | (0.06) | (0.04) |
| somecllg | -0.000 | -0.000 | -0.031*** |
| | (0.01) | (0.01) | (0.01) |
| baab | -0.004 | 0.018*** | -0.005 |
| | (0.00) | (0.01) | (0.00) |
| lnstem | 0.034* | -0.023 | 0.042*** |
| | (0.02) | (0.02) | (0.01) |
| lnctyemp | 0.049 | 0.272*** | 0.128*** |
| | (0.05) | (0.07) | (0.04) |
| prime | 0.033*** | 0.020 | 0.032*** |
| | (0.01) | (0.02) | (0.01) |
| htempr | 0.003 | 0.008 | -0.008 |
| | (0.01) | (0.01) | (0.01) |
| unempr | 0.026** | 0.189*** | 0.023** |
| | (0.01) | (0.02) | (0.01) |
| vc_bin | -0.667*** | -0.642** | -0.744*** |
| | (0.21) | (0.30) | (0.20) |

| | FDI employment | FDI received or not | No. of projects |
|---|---|---|---|
| | (Model 1) | (Model 2) | (Model 3) |
| lnvc | 0.041*** | 0.046** | 0.043*** |
| | (0.01) | (0.02) | (0.01) |
| lnkspl_50 | -0.066 | 0.169*** | -0.168*** |
| | (0.05) | (0.06) | (0.04) |
| meantravel | 0.019** | -0.049*** | 0.013** |
| | (0.01) | (0.01) | (0.01) |
| lnlnmil | 0.001 | 0.287** | -0.108** |
| | (0.07) | (0.12) | (0.05) |
| ap | -0.001 | 0.004 | 0.004 |
| | (0.00) | (0.00) | (0.00) |
| lnelectr | -0.256** | -0.140 | -0.217** |
| | (0.13) | (0.18) | (0.11) |
| Buzclmt15 | 0.008 | 0.038 | 0.058* |
| | (0.03) | (0.04) | (0.03) |
| r2w_flag | 0.058 | 0.055 | 0.058 |
| | (0.06) | (0.09) | (0.04) |
| tb1 | -0.000 | -0.000* | -0.000** |
| | (0.00) | (0.00) | (0.00) |
| tb2 | -0.005* | -0.000 | -0.000 |
| | (0.00) | (0.00) | (0.00) |
| ppf | -0.002 | -0.004 | -0.002* |
| | (0.00) | (0.00) | (0.00) |
| credit | 0.030 | 0.151*** | 0.037* |
| | (0.02) | (0.03) | (0.02) |
| Constant | 0.943 | -14.242*** | -3.238*** |
| | (0.93) | (1.33) | (0.72) |
| ln(α) | | | -1.234*** |
| | | | (0.09) |
| Observations | 5,032 | 55,635 | 5,033 |
| Adj. $R^2$ | 0.159 | | |
| Pseudo $R^2$ | | 0.220 | 0.173 |

Notes: FDI employment was estimated by OLS, the binary FDI logit model and no. of projects negative Binomial model. The significance of *a* coefficient suggests that the data is highly skewed and in support of using the negative Binomial model. Note that the sample size is greatly reduced, because for employment and project counts, we only restrict our sample to positive values. Significance: * 10%, ** 5%, *** 1%.

## Table A2: Summary of Empirical Results of Panel Models for FDI Employment

| | Panel I | | | Panel II | |
|---|---|---|---|---|---|
| | WE | BE | RE | WE | BE |
| lnclstemp | 0.287 | 0.149*** | 0.180*** | 0.201*** | 0.191*** |
| | (0.31) | (0.02) | (0.02) | (0.04) | (0.04) |
| sei_clst | -5.186 | -0.309 | -0.407 | *o.m.* | -1.521 |
| | (3.62) | (0.54) | (0.51) | | (1.01) |
| lnclst_bal1 | -0.009 | 0.122*** | 0.133*** | 0.290*** | 0.118*** |
| | (0.11) | (0.02) | (0.02) | (0.05) | (0.04) |
| htflag | *o.m.* | 0.667 | 1.142* | 2.990** | -0.275 |
| | | (0.68) | (0.62) | (1.44) | (1.83) |
| lq_bin | -0.054 | 0.112** | 0.119** | 0.077 | -0.033 |
| | (0.14) | (0.05) | (0.05) | (0.11) | (0.14) |
| lnlq_clst | -0.136 | -0.048* | -0.060** | 0.014 | -0.041 |
| | (0.33) | (0.03) | (0.03) | (0.06) | (0.08) |
| lq_bin lnlq_clst | -0.037 | 0.136*** | 0.154*** | 0.240** | 0.174* |
| | (0.24) | (0.04) | (0.04) | (0.09) | (0.09) |
| htflag x lnclstemp | -0.333** | -0.037** | -0.047*** | -0.098*** | -0.170*** |
| | (0.16) | (0.02) | (0.02) | (0.03) | (0.06) |
| htflag x sei_clst | -6.376 | -0.436 | -0.931 | -2.749 | 1.680 |
| | (5.31) | (0.85) | (0.78) | (1.83) | (2.20) |
| htflag x lnclst_bal1 | 0.031 | -0.145*** | -0.159*** | -0.383*** | -0.262** |
| | (0.19) | (0.04) | (0.04) | (0.09) | (0.11) |
| coli_trsp | *o.m.* | -0.005* | -0.002 | *o.m.* | -0.017** |
| | | (0.00) | (0.01) | | (0.01) |
| nohs | -0.187 | -0.021 | -0.038 | *o.m.* | -0.066 |
| | (0.12) | (0.03) | (0.03) | | (0.06) |
| somecllg | 0.082** | -0.002 | -0.004 | *o.m.* | 0.008 |
| | (0.04) | (0.00) | (0.01) | | (0.01) |
| baab | 0.061* | -0.002 | -0.001 | *o.m.* | -0.007 |
| | (0.04) | (0.00) | (0.00) | | (0.01) |
| lnstem | 0.079 | -0.007 | -0.007 | *o.m.* | 0.002 |
| | (0.06) | (0.01) | (0.01) | | (0.01) |
| prime | -0.114*** | 0.030*** | 0.031*** | *o.m.* | 0.027* |
| | (0.04) | (0.01) | (0.01) | | (0.01) |
| htempr | -0.039 | -0.001 | -0.005 | *o.m.* | 0.013 |
| | (0.06) | (0.01) | (0.01) | | (0.01) |
| unempr | 0.110*** | 0.024*** | 0.042*** | *o.m.* | 0.074*** |
| | (0.02) | (0.01) | (0.01) | | (0.03) |
| vc_bin | -0.106 | -0.807*** | -0.580*** | *o.m.* | -1.371*** |
| | (0.41) | (0.20) | (0.17) | | (0.50) |
| lnvc | 0.008 | 0.050*** | 0.033*** | *o.m.* | 0.087*** |
| | (0.03) | (0.01) | (0.01) | | (0.03) |
| lnkspl_50 | | | | | 0.134** |
| | | | | | (0.07) |

| | Panel I | | | Panel II | |
|---|---|---|---|---|---|
| | WE | BE | RE | WE | BE |
| meantravel | -0.016 | 0.012*** | 0.011** | *o.m.* | -0.003 |
| | (0.06) | (0.00) | (0.00) | | (0.01) |
| lnlnmil | | 0.078** | 0.092** | *o.m.* | 0.076 |
| | | (0.03) | (0.05) | | (0.08) |
| ap | 0.023 | -0.001 | 0.004 | *o.m.* | -0.001 |
| | (0.02) | (0.00) | (0.00) | | (0.01) |
| lnelectr | -0.368 | -0.185** | -0.220** | *o.m.* | -0.074 |
| | (0.31) | (0.09) | (0.11) | | (0.19) |
| buzclmt15 | | 0.003 | -0.002 | *o.m.* | 0.019 |
| | | (0.03) | (0.03) | | (0.05) |
| r2w_flag | *o.m.* | 0.085** | 0.099** | *o.m.* | 0.063 |
| | | (0.04) | (0.05) | | (0.08) |
| tb1 | *o.m.* | -0.000 | -0.000 | *o.m.* | -0.000 |
| | | (0.00) | (0.00) | | (0.00) |
| tb2 | *o.m.* | -0.003 | -0.005** | *o.m.* | -0.010** |
| | | (0.00) | (0.00) | | (0.00) |
| ppf | *o.m.* | 0.001 | 0.001 | *o.m.* | 0.002 |
| | | (0.00) | (0.00) | | (0.00) |
| credit | 0.013 | 0.007 | 0.015 | *o.m.* | -0.018 |
| | (0.05) | (0.01) | (0.01) | | (0.03) |
| Constant | 8.515** | 2.601*** | 2.102*** | 2.715*** | 4.797*** |
| | (4.27) | (0.70) | (0.68) | (0.27) | (1.40) |
| Observations | 8,171 | 8,171 | 8,171 | 2,434 | 2,434 |
| Within $R^2$ | 0.061 | 0.000 | 0.005 | 0.097 | 0.040 |
| Between $R^2$ | 0.000 | 0.099 | 0.097 | 0.086 | 0.195 |
| Overall $R^2$ | 0.001 | 0.126 | 0.128 | 0.089 | 0.112 |

Notes: The test statistic of over-identifying restriction on the RE model in Panel I is $\chi^2$ (18)=129.89 that rejects the null that is in favor of the FE models. Panel II (the cluster panel) used only the sample of last time period (2012-2015) and estimated FE models, due to unbalanced panel. Statistical significance: *10%, ** 5% and *** 1%.

## Table A3: Variable Names and Definitions

| Variable name | Definition |
|---|---|
| lnclstemp | Cluster employment, county, log |
| sei_clst | Between cluster Shannon Evenness Index, county |
| lnclst_bal1 | Within-cluster balance, log |
| htflag | High-tech industry (cluster) flag, binary |
| lq_bin | Cluster location quotient above 1.2 value, binary |
| lnlq_clst | Cluster location quotient, log |
| lq_bin x lnlq_clst | Interaction: Cluster LQ above 1.2 (bin) and cluster LQ (log) |
| htflag x lnclstemp | Interaction: High-tech industry flag and cluster employment |
| htflag x sei_clst | Interaction: High-tech industry flag and between cluster evenness |
| htflag x lnclst_bal1 | Interaction: High-tech industry flag and within-cluster industry balance (log) |
| coli_trsp | Cost of living index for transportation, county |
| nohs | Educational attainment: no high school, county, proportion |
| somecllg | Educational attainment: some college, county, proportion |
| baab | Educational attainment: bachelor's degree, county, proportion |
| lnstem | STEM occupations, county, proportion, log |
| prime | Population between 25 and 45 years of age, county, proportion |
| htempr | Employment in high-tech industries, county, proportion |
| unempr | Unemployment rate, county |
| vc_bin | Venture capital, county, binary |
| lnvc | Venture capital, county, dollar value, log |
| lnkspl_50 | Knowledge spillovers from universities within 50 miles, county, log |
| meantravel | Mean communting time to work, county |
| lnlnmil | Interstate lane miles per capita, county, log |
| ap | Average wage in manufacturing, state, dollars per hour |
| lnelectr | Electricity rates, state |
| buzclmt15 | Business climate, state |
| r2w_flag | Right to work, state |
| tb1 | Tax burden, state, dollars per capita |
| tb2 | Tax burden, state, percent of state income |
| ppf | Public pension plans, percent funded, state |
| credit | Credit worthiness, state |